Wilson KJ, Heaps SE, Farrow M.

[Demand forecasting over complex geographical networks: the case of Northern Gas Networks](#).

***In: 26th European Safety and Reliability Conference*. 2016, Glasgow: CRC Press.**

Demand forecasting over complex geographical networks: the case of Northern Gas Networks

K.J. Wilson, S.E. Heaps & M. Farrow
School of Mathematics and Statistics, Newcastle University, UK

ABSTRACT: Organisations which distribute resources over complex networks need to be able to meet demand. They require accurate forecasts of demand across the network. We consider short term demand forecasting for Northern Gas Networks (NGN), who deliver gas to 2.7 million homes and businesses. They require forecasts at different locations for each hour over each day. The locations are of distinct types by the dominant user groups. For each we develop a time series model to forecast short term demand, utilising seasonality at monthly, daily and hourly levels and temperature. Annually NGN produce Gas Demand forecasts for a decade, which support investment and planning and inform National Grid. We develop time-series models and Bayesian inference to provide forecasts of daily demand. A challenge arises in modelling demand by industry where weather has less predictive power and step changes occur due to large consumers switching to or from gas.

1 INTRODUCTION

The Gas distribution system in the UK is complex, with the four main stages of production/importation, transmission, distribution and supply being the responsibility of different organisations. Northern Gas Networks (NGN) is one of 8 regional distribution networks who take gas from the national grid and transport it for delivery to customers. Thus it has a vital role in maintaining a reliable gas supply to customers. Therefore, it requires accurate short term forecasts of demand at individual points in the network and long term forecasts of daily gas demand over larger areas known as Local Distribution Zones (LDZs). That is, separate forecasting is used for balancing supply and demand at a local level and for long term decision making.

Within the regional network, there are a number of sites which take off gas to supply locally. These are known as offtakes. For the first objective, NGN forecast the gas demand at each pressure controlled offtake for each hour over each 24 hour period. There are additional volumetric offtakes which are not part of this work. For the second objective, NGN produces gas demand forecasts for each day over a ten year period.

The 2.7 million homes and businesses supplied by NGN are divided into two LDZs; North and North East. Currently hourly demand over the next 24 hours is forecast for each LDZ using a profile from a day in a previous year which is similar to the current day (temperature, month, day of week) and this is scaled down for each offtake based on their share of the gas demand up to the present time. Also, for each LDZ, daily forecasts are made. These forecasts are then adjusted using a scale factor for weekends and holiday periods and are fed into a larger forecasting process which takes into account changes in the price of oil, government policy and so on. The daily forecasts are made for 5 different load bands; 0-73MWh (e.g. a single house), 73-732MWh (e.g. a large block of flats), 732-2196MWh, 2196-5860MWh and > 5860MWh (various sizes of industrial premises). For more information on current approaches in the UK see National Grid (2012). Both the long term daily forecasts and the short term hourly forecasts could benefit from careful time series modelling.

Typical approaches to time series modelling fall into two broad categories, auto-regressive integrated moving average (ARIMA) models and dynamic linear models (DLM) (Shumway & Stoffer 2011). In both cases, models are built using seasonal components, trends and covariate information, typically in a linear Gaussian fashion. DLMs tend to be more interpretable in terms of model building, allowing different components such as seasonality to be built separately before being combined. ARIMA models have advantages in terms of inference, although DLMs can always be converted into ARIMA models for this purpose.

In this paper we build models to forecast short term hourly gas demand for individual offtakes and long term daily gas demand over an LDZ. In both cases, we model the natural logarithm of demand. We consider daily and yearly seasonal components, modelled as Fourier series, and a covariate effect of weather. For the hourly forecasts, we also include an effect for hour of the day and in the daily case, we allow for the effects of public holidays. We include both additive and multiplicative effects. We take a Bayesian approach to inference and so thought is given to appropriate prior specification. For more information on applied Bayesian time series modelling see, for example, Chapter 8 of Congdon (2006).

Sections 2–4 focus mainly on the short-term hourly forecasting problem. In Section 2 we perform an initial exploratory analysis of demand data from NGN. Section 3 then gives the general form of our time series model and the specific form for the hourly forecasting. In Section 4 we show the results of the modelling for the NGN data. Finally, Section 5 considers the long-term daily forecasting problem, Section 6 discusses the anticipated impact of the modelling for NGN and in Section 7 we summarise the main points of the paper.

2 EXPLORATORY ANALYSIS

In this section, for reasons of space, we focus solely on hourly demand in the Northern LDZ. Demand is given over an hour in millions of cubic metres. Based on the expert judgement of NGN, the following potential influences in the demand for gas have been identified: annual seasonality, seasonality by the day of the week, seasonality by the hour of the day and a covariate effect of temperature.

We plot the natural logarithm of gas demand by each offtake in 2012 in Figure 1.

There are some important messages to be drawn from the figure. Firstly, there is evidence of clear seasonal behaviour across the year in many of the time series. Offtake demand is higher at the beginning and end of the year than in the middle. This seasonal behaviour is unlikely to be independent of weather, however. Secondly, there are clear differences between the pattern in offtake demand between the different offtakes. We will return to this point later.

If we reduce the timescale on the x-axis on our time series plot then this will allow us to investigate the possibility of seasonality in the data by day of the week and hour of the day. In Figure 2 we give time series plots of each of the offtake demands for the offtakes in the Northern region in week 3 (15th-22nd) of January 2012.

We see a daily pattern in gas offtake demand, with two clear peaks around 8am and 6pm visible in most of the time series. There is clear evidence of seasonality by hour of the day in many of the series. We also see evidence of differences by day of the week, with the 15th January (a Sunday) and the 21st of January (a Saturday) typically having a less pronounced pattern than the other days in the week, the 16th-19th (Monday-Thursday) having similar patterns of offtake

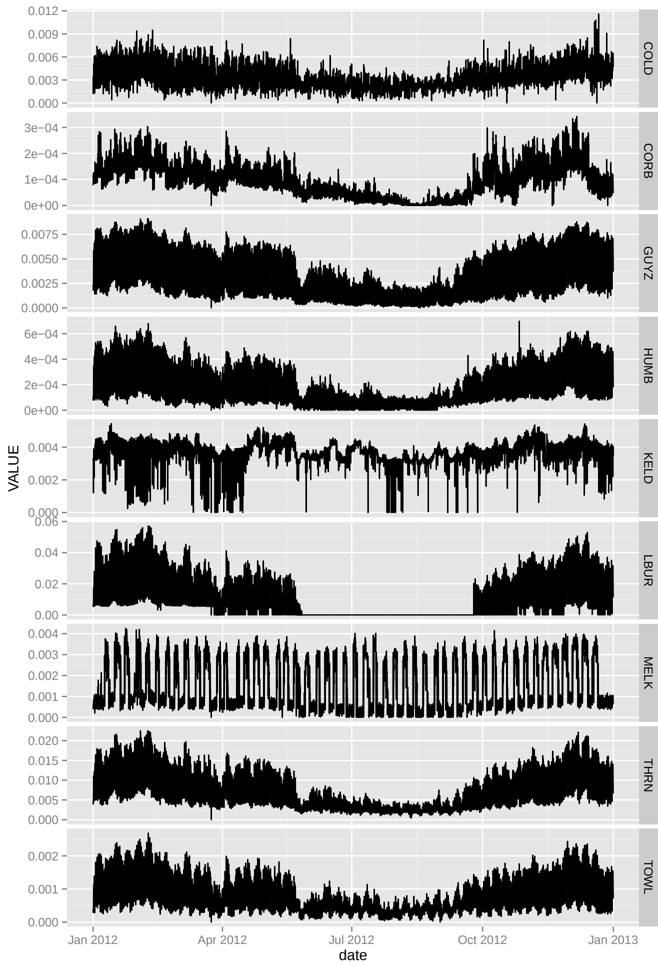

Figure 1: The offtake demand for the offtakes in the Northern LDZ for each hour over 2012.

demand and a differing pattern on the 20th (Friday). However, there are also large differences between individual offtakes, with some, such as Tow Law, having a clear daily profile and others, such as Keld, looking more random in nature.

We can also consider the potential covariate effect of weather, in this case in terms of temperature. To investigate this, we choose a single hour of the day, 9am, and produce scatter plots of temperature against offtake demand. They are given in Figure 3.

We see reasonably linear relationships between the two variables in this case, particularly at low- to mid-range temperatures. However, there is some evidence of non-linearity for higher temperatures, for example in the plot for Thrintoft. In residential areas, it could be the case that there is virtually zero demand for gas on reasonably hot days, which is causing this effect.

We see that the well-behaved offtake demand time series in the Northern LDZ are Coldstream, Guyzance, Humbleton, Thrintoft and Tow Law. These offtakes are dominated by domestic demand. The more complex offtake demand time series in the Northern LDZ are Corbridge, Keld, Little Burdon and Melkinthorpe.

Of the more complex patterns, Little Burdon has long spells during which there is zero demand. This is because this offtake is turned off during the summer and the demand is transferred to a volumetric off-

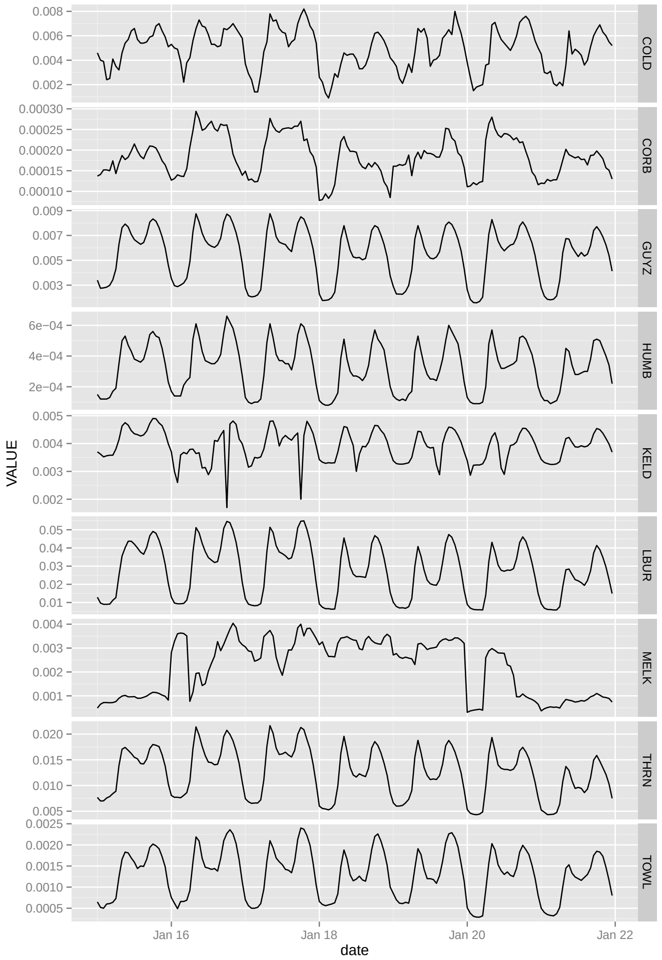

Figure 2: The offtake demand for the offtakes in the Northern LDZ for each hour over the third week of January 2012.

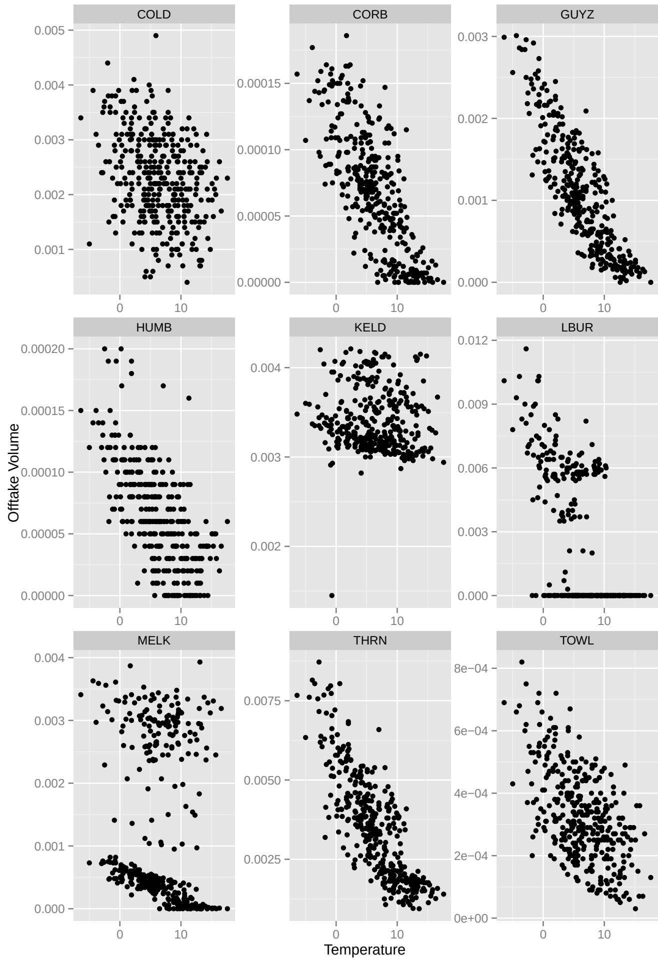

Figure 3: The temperature against offtake demand volume for the offtakes in the Northern LDZ at 9am for every day in 2012.

take. The rest of the time the offtake demand at Little Burdon is well-behaved. Corbridge supplies an army barracks. This produces different offtake demand profiles when the army is present at the barracks or away. Melkinthorpe is home to both residential users of gas and the industrial organisation British Gypsum. The offtake demand pattern will be heavily influenced by British Gypsum's industrial process schedules. Keld similarly has a mix of residential and industrial users.

3 TIME SERIES MODELLING

The variable of interest is the gas demand in n locations at time t, $\boldsymbol{D}_t = (D_{t,1}, \ldots, D_{t,n})$. As well as seasonal components, this will depend on a continuous weather variable (e.g. temperature) measured at m locations, $\boldsymbol{W}_t = (W_{t,1}, \ldots, W_{t,m})$. We model the natural logarithm of demand,

$$\boldsymbol{Y}_t = \log(\boldsymbol{D}_t).$$

We construct a joint first order autoregressive (AR(1)) model for the log demand and the weather variable at time t. This can be factorised as

$$\begin{aligned} & p(\boldsymbol{y}_t, \boldsymbol{w}_t \mid \boldsymbol{y}_{t-1}, \boldsymbol{w}_{t-1}, \boldsymbol{\theta}) \\ & = p(\boldsymbol{y}_t \mid \boldsymbol{w}_t, \boldsymbol{y}_{t-1}, \boldsymbol{w}_{t-1}, \boldsymbol{\theta}) p(\boldsymbol{w}_t \mid \boldsymbol{w}_{t-1}, \boldsymbol{\theta}) \\ & = p(\boldsymbol{y}_t \mid \boldsymbol{w}_t, \boldsymbol{\theta}^y) p(\boldsymbol{w}_t \mid \boldsymbol{w}_{t-1}, \boldsymbol{\theta}^w), \end{aligned}$$

where $p(\cdot)$ represents a density function, $\boldsymbol{\theta}$ is a vector of model parameters and $(\boldsymbol{\theta}^y, \boldsymbol{\theta}^w)$ are the parameters associated with the conditional model for $\boldsymbol{y}_t$ given $\boldsymbol{w}_t$ and the marginal model for the $\boldsymbol{w}_t$ respectively. Thus we see that the demand depends on previous time points only through its relationship with the weather variable.

The conditional model for demand is given by

$$\boldsymbol{Y}_t \mid \boldsymbol{W}_t = \boldsymbol{w}_t, \boldsymbol{\theta}^y \sim \mathrm{N}_n(\boldsymbol{\mu}_t, \Omega^{-1}),$$

where Ω is a static precision matrix and the conditional mean $\boldsymbol{\mu}_t = (\mu_{t,1}, \ldots, \mu_{t,n})$ is given by

$$\mu_{t,j} = \log(|\alpha_{t,j}|) + \beta_{t,j}, \tag{1}$$

for $j = 1, \ldots, n$, where $\alpha_{t,j}$ and $\beta_{t,j}$ represent multiplicative and additive effects on demand respectively and can depend on both the effect of weather and other covariates and seasonality at different time scales. The modulus is used to ensure that the logarithm has a value, though in the applications in this paper $\alpha_{j,t}$ has always been positive and so the modulus has not been needed.

We suppose the weather variable represents some change from the average at that particular time (e.g. day of the year). In this case, a suitable model for the weather variable at time t is of the form

$$\boldsymbol{W}_t \mid \boldsymbol{W}_{t-1}, \boldsymbol{\theta}^w \sim \mathrm{N}_m(\Phi \boldsymbol{w}_{t-1}, \Psi^{-1}),$$

where Φ is an $m \times m$ matrix and Ψ is a static precision matrix.

3.1 *Hourly demand*

We are interested in forecasting the demand at each of the offtakes individually. For illustration in this paper, we just consider the $n = 6$ well behaved offtakes in the Northern LDZ identified in Section 2. The weather data available is temperature and it is given at the LDZ level ($m = 1$), so for the Northern LDZ the temperature at time t is $\tilde{X}_t$, where time is measured in hours. The weather variable to be included in the model is then

$$W_t = \tilde{X}_t - m_{X,t},$$

where $m_{X,t}$ is the mean temperature for the LDZ in that hour of the year based on historical data. Thus W_t is the change from the average temperature in the LDZ in that hour of the year.

The specific form of the model for hourly forecasting of the individual offtakes is

$$\begin{aligned} \alpha_{t,j} &= \xi_{y(t),j} + \lambda_j x_t, \\ \beta_{t,j} &= \zeta_{y(t),j} + \tau_{t,j} + \chi_{y(t),t,j}, \end{aligned}$$

where $y(t)$ represents the day of the year associated with time t, $\xi_{y(t),j}$ represents a seasonal effect with regards to time of the year, λ_j represents the effect of temperature, $\zeta_{y(t),j}$ represents a day of the week effect, $\tau_{t,j}$ is an hour of the day effect and $\chi_{y(t),t,j}$ is an interaction effect between the hour of the day and the day of the week.

We would like to incorporate an annual cycle in the model in a parsimonious way. A suitable form for these effects is therefore a truncated Fourier series of the form

$$\xi_{t,j} = \gamma_{j,0} + \sum_{k=1}^{K} \left\{ \gamma_{j,1,k} \cos\left(\frac{2\pi kt}{365.25}\right) + \gamma_{j,2,k} \sin\left(\frac{2\pi kt}{365.25}\right) \right\},$$

where $K \leq 182$ controls the number of harmonics of the series and hence the number of parameters representing seasonal effects. In the case of $K = 2$, for example, we see that we would have 5 parameters to represent seasonal effects across the year.

Similarly, we can represent the day of the week effects in terms of a Fourier series. In this case, a suitable form is

$$\zeta_{t,j} = \sum_{k=1}^{3} \left\{ \delta_{j,1,k} \cos\left(\frac{2\pi kt}{7}\right) + \delta_{j,2,k} \sin\left(\frac{2\pi kt}{7}\right) \right\},$$

and so we have the correct number of free parameters, 6, to represent the effects of the 7 days of the week as fixed effects.

The hour of the day effect can also be expressed as a Fourier series in the following manner,

$$\tau_{t,j} = \sum_{k=1}^{11} \left\{ \eta_{j,1,k} \cos\left(\frac{2\pi kt}{24}\right) + \eta_{j,2,k} \sin\left(\frac{2\pi kt}{24}\right) \right\} + \eta_{j,1,12} \cos(\pi t),$$

which has the correct number of free parameters, 23, to represent each hour of the day.

In order to perform inference for the model, we are required to specify prior distributions for all of the model parameters. In this case, the conjugate prior distribution for the precision, Ω, is a Wishart distribution,

$$\Omega \sim \mathrm{W}_6(I_6, \nu_\Omega),$$

and, as Ψ is only one-dimensional, an appropriate conjugate prior distribution for Ψ is

$$\Psi \sim \text{gamma}(a_\Psi, b_\Psi),$$

for constants a_Ψ, b_Ψ.

In terms of the autoregressive coefficient Φ, a stationary process is suitable for the change in temperature compared to the average and so we require $|\Phi| < 1$. Also, we would expect an above average temperature in the previous hour to have a positive effect on the temperature in the next hour. Therefore a suitable prior distribution is

$$\Phi \sim \text{beta}(c_\Phi, d_\Phi),$$

for constants c_Φ, d_Φ.

For each of the seasonal parameters, we consider priors which relate the different offtakes as we would expect the seasonal effects to be similar across the offtakes. Therefore we choose a prior which borrows strength across the six offtakes. That is, for seasonal effect S_j, we define a prior distribution

$$S_j \sim \mathrm{N}(\mu_{S,j}, (1 - r_S) v_S), \quad (2)$$

$$\mu_{S,j} \sim \mathrm{N}(0, r_S v_S), \quad (3)$$

where (r_S, v_S) are the correlation between offtakes and offtake marginal variance respectively. In practice, we choose each $r_S = 0.85$ and v_S takes the value 100. The prior distribution for λ_j also takes this form.

3.2 *Extensions to the model*

There are extensions to the model to incorporate forecasting for the more complex offtakes. Space does not permit a detailed description here. A brief description is given below.

In the hourly forecasting, there are four offtakes in which demand does not behave in the same way as the others. Little Burdon operates in a similar pattern to the well behaved offtakes over the winter months, with the typical two-peaked shape each day and the day of the week effects, but is switched off for the summer months when gas demand is lower. We can define an indicator variable which takes the value 1 when Little Burdon is operating and 0 when it is switched off. When it is operating it follows the standard demand model. In the case of Melkinthorpe, a two-component mixture model is suitable, with one component representing domestic users and taking the form of the model in the previous section and the other representing industrial users, primarily British Gypsum. Keld also has a mix of domestic and industrial users and so a mixture model can also be used.

4 RESULTS

We consider hourly demand data at the offtake level and two-hourly temperature data at the LDZ level from January, February and March 2012. Initially, we wish to see if the model is sufficiently flexible to capture the general hourly behaviour of the demand in each offtake as displayed in Figures (1) and (2). Therefore we fit the model given in the previous section to all of the data and, for each offtake, find the posterior mean and 95% posterior credible intervals for $\mu_{t,j}$, the mean log demand. The results for a single month are given in Figure 4.

We see different profiles across the different offtakes, with Corbridge in particular showing a different demand pattern to the other 5 offtakes. This is consistent with the observations made in the exploratory analysis. We also see deeper troughs of demand in Guyzance in comparison to the other offtakes, Coldstream showing different behaviour to the other offtakes and the profiles for Humbleton, Thrintoft and Tow Law looking fairly similar. We see variation across the month in the daily patterns in most of the profiles, as a result of the covariate effect of temperature and the weekly seasonal effects.

Of course, the model is designed to forecast demand in the future. Therefore, we need to look at forecasts from the model for days in which we do not have data. We consider the same period of data to which we fit the model before, but now exclude the final week of observations. We then predict these demand values using the fitted model. We report posterior predictive means and 95% posterior predictive intervals. The results are given in Figure 5 for the 6 well-behaved offtakes. Also plotted are the observed log demand values.

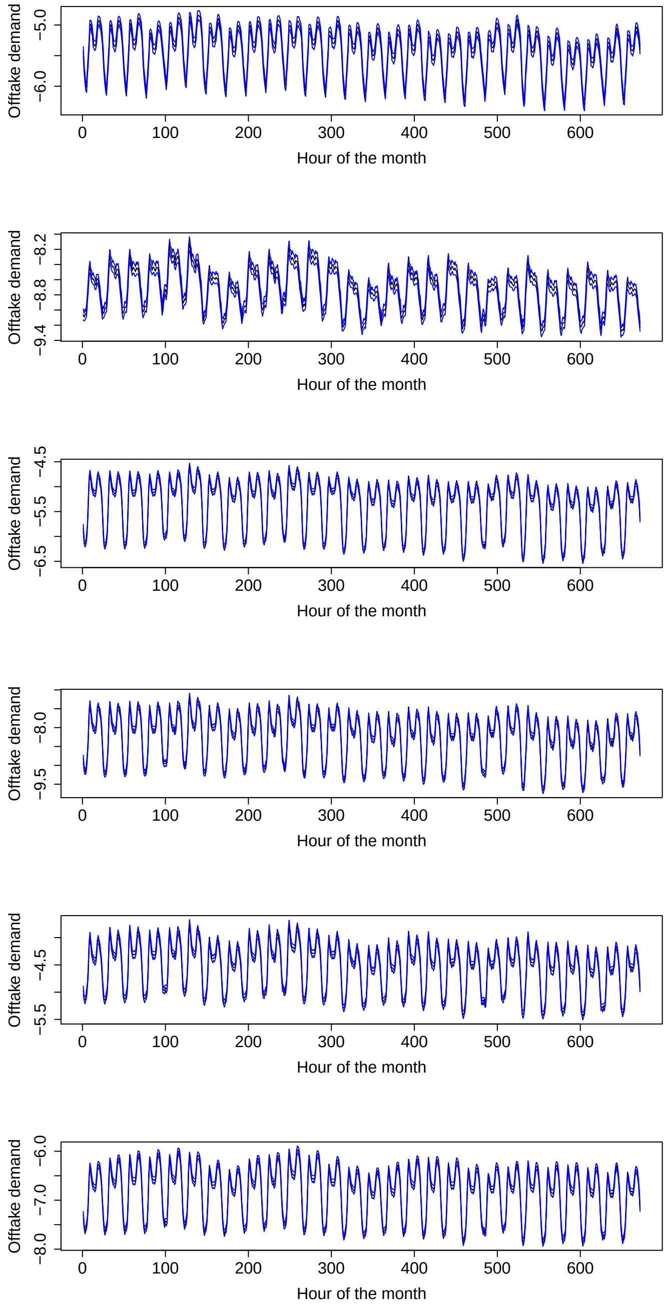

Figure 4: Posterior means (black) and 95% credible intervals (blue) for the mean natural logarithm of demand for the six well-behaved offtakes in the Northern LDZ. From top to bottom they are Coldstream, Corbridge, Guyzance, Humbleton, Thrintoft and Tow Law.

We see that the model is capturing the general shapes of the daily profiles well for each offtake. The posterior predictive intervals contain the true observation the majority of the time. The model is not capturing the effect of high temperatures particularly well, however, typically overestimating the demand for these values. One possiblity to extend the model to overcome this is to incorporate explicit time dependence into the demand forecasts, for example as an autoregressive process as with temperature. Of course, it is most important to meet demand in high-demand periods and this is where the model is particularly strong.

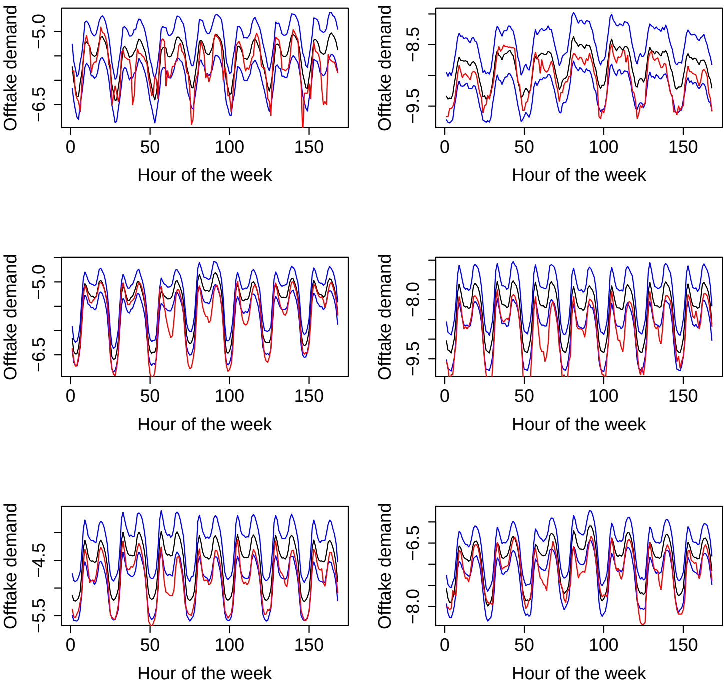

Figure 5: Posterior predictive means (black) and 95% posterior prediction intervals (blue) for the natural logarithm of demand for the six well-behaved offtakes in the Northern LDZ. Also plotted are the observed log demand values (red). From top left to bottom right they are Coldstream, Corbridge, Guyzance, Humbleton, Thrintoft and Tow Law.

5 FORECASTING DAILY DEMAND

5.1 *The Model*

We can model the daily demand in much the same way as hourly demand, with a few small changes. In this case, a separate forecast is made for each of 4 different load bands; 0-73MWh, 73-732MWh, 732-2196MWh and 2196-5860MWh. That is, a model of the form of those in Section 3 is adopted for each load band in turn. In this case, the log demand is $(Y_{t,1}, Y_{t,2})$, representing the demand in the Northern and North Eastern LDZ respectively.

The weather variable used for daily forecasting is a construct used by Northern Gas Networks called the "composite weather variable". This is constructed to have a linear relationship with gas demand. Suppose that the composite weather variable for LDZ j on day t is $\tilde{W}_{t,j}$. Then the covariate for weather used in the model is

$$W_{t,j} = \tilde{W}_{t,j} - m_{W,j},$$

where $m_{W,j}$ is the mean value of the composite weather variable on that day from historical data. The model for daily demand takes the form of (1) with $\alpha_{t,j}$ and $\beta_{t,j}$ given by

$$\alpha_{t,j} = \xi_{t,j} + \lambda_j w_{t,j},$$

$$\beta_{t,j} = \zeta_{t,j} + \sum_{k=1}^{4} \beta_{j,k} x_{t,k},$$

where $(\xi_{t,j}, \lambda_j, \zeta_{t,j})$ are defined as in the hourly model, $\beta_{j,k}$ are public holiday effects and $\boldsymbol{x}_t = (x_{t,1}, x_{t,2}, x_{t,3}, x_{t,4})$ are indicator variables which represent Easter public holidays (Good Friday, Easter Monday), Christmas public holidays (Christmas Day, Boxing Day, New Year's Day), other public holidays (May Day, Spring Bank Holiday, etc.) and days in proximity to a public holiday (typically the weekend immediately before or after the public holiday).

The seasonal and day of the week effects, $\xi_{t,j}$ and $\zeta_{t,j}$, take the form of Fourier series as in the hourly model.

In order to perform inference for the model, we are required to specify prior distributions for all of the model parameters. In the case of the precision matrices Ω and Ψ, suitable conjugate prior distributions are Wishart distributions of the form

$$\Omega \sim \mathrm{W}_2(I_2, \nu_\Omega), \quad \Psi \sim \mathrm{W}_2(I_2, \nu_\Psi),$$

where I_2 is the (2×2) identity matrix. In practice, we choose $\nu_\Omega = \nu_\Psi$ to be 3.

Although a stationary process would, again, seem sensible for the weather variable, for convenience we adopt a prior distribution for the coefficients of the autoregressive process Φ of the form

$$\Phi \sim \mathrm{N}_2(0, CI_2),$$

for some constant C. The value chosen in this case is $C = 100$. We note that, in the applications to daily data, the posterior mass is concentrated in the stationarity region.

For the seasonal effects and covariate effects, we adopt normal priors which borrow strength across different LDZs of the form of (2) and (3).

We use a similar structure for the public holiday effects. In this case, we consider the vector $\boldsymbol{\beta}_j = (\beta_{j,1}, \ldots, \beta_{j,4})$ and assume a prior of the form

$$\boldsymbol{\beta}_j \sim \mathrm{N}_4(\mu_{\beta,j}, (1 - r_\beta)\boldsymbol{V}_\beta),$$

$$\boldsymbol{\mu}_{\beta,j} \sim \mathrm{N}(\boldsymbol{0}_4, r_\beta \boldsymbol{V}_\beta),$$

in which $r_\beta = 0.85$ and $\boldsymbol{V}_\beta$ is a 4×4 matrix specified with non-zero off-diagonal elements to represent beliefs that the effects of different kinds of public holidays are likely to be similar.

5.2 *Extensions to the model*

There is a fifth load band (> 5860MWh) corresponding to big industrial users which behaves very differently to the other four. In particular, the same seasonality across the year is not evident in load band 5 and so the model can be simplified. In the Northern LDZ there appears to be a piecewise linear trend in the time series and a Markov chain can be used to move from one segment to another. In the North East LDZ there is evidence of seasonality and step changes in the average level of demand. A hidden Markov model can be used to model this (Rabiner & Juang 1986).

5.3 *Results*

We test the model by fitting it to a subset of data, omitting one year, and making predictions for the missing year. The point predictions behave well in comparison to those produced by the existing methodology in NGN (Figure 6), with the added advantage that we are able to produce prediction intervals and observe that the actual values fall within the prediction intervals for approximately the predicted proportion of the time.

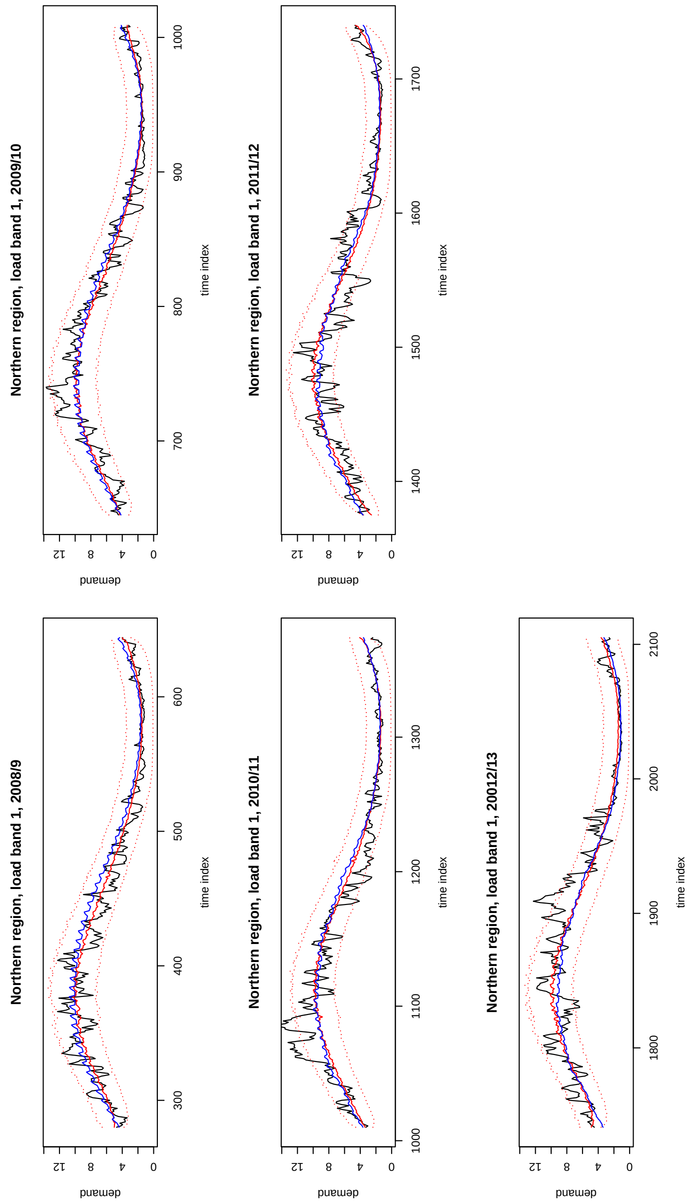

Figure 6: Comparisons of predictions, Northern region. Demand values are divided by 10^4. Each year starts in October. Time index 1 corresponds to 1st January 2008. The black line is observed demand, the blue line is NGN's old methodology forecast, the red solid line is the posterior predictive mean and the red dotted lines are 2.5% and 97.5% points in the posterior predictive distribution.

The figure shows "leave one year out" predictions for Load Band 1. The solid red lines are posterior predictive means, the dotted red lines are 95% prediction intervals, the black line is the actual observed demand and the blue line is the point prediction supplied by NGN using the old methodology. We see that the new approach is capturing the uncertainty in the forecasts in a way which was not previously available to NGN.

The methodology developed also allows us to estimate the day of the week effects in a systematic way. A subset of the posterior means and 95% credible intervals for the daily effects in Load Band 2 are given in Table 1.

	Northern LDZ	
Day	Load Band 2	
Monday	1.124	(1.111,1.137)
Tuesday	1.122	(1.109,1.134)
Wednesday	1.124	(1.111,1.137)
Thursday	1.108	(1.095,1.121)
Friday	1.050	(1.038,1.061)
Saturday	0.759	(0.750,0.768)
Sunday	0.800	(0.791,0.809)
	North Eastern LDZ	
Day	Load Band 2	
Monday	1.121	(1.106,1.137)
Tuesday	1.126	(1.111,1.142)
Wednesday	1.124	(1.109,1.140)
Thursday	1.114	(1.099,1.129)
Friday	1.053	(1.039,1.068)
Saturday	0.776	(0.766,0.788)
Sunday	0.774	(0.763,0.768)

Table 1: Multiplicative effect of each day of the week in the two LDZs. Estimates are posterior means with 95% equi-tailed credible intervals.

We see that the weekdays Monday-Thursday are reasonably interchangeable in both LDZs, with gas demand being high on these days as indicated by the effects being greater than 1. Friday also has an effect larger than 1 indicating that demand is high, although it is smaller in both LDZs than those for Monday-Thursday. By contrast, Saturday and Sunday have effects smaller than 1, indicating that demand for gas is relatively low on these days.

6 DISCUSSION

At present, regular, significant human interventions are required from NGN to maintain gas supply balance. Better utilisation of historical data through the modelling described in this paper will provide a robust, systemised statistical approach to forecasting, increasing the accuracy of offtake gas demand forecasts and thereby reducing the number of corrective (human intervention) actions required to maintain an accurate gas supply balance and minimise the errors in capacity profile submissions to the National Grid.

The methodology will be incorporated into Northern Gas Networks' Offtake Profile forecasting software system so that submissions to National Grid can be provided with greater accuracy. The Health and Safety Executive (HSE) like to see evidence of safety focus and attention excellence within gas networks and high quality management of the gas demand systems process can help to achieve this.

A key advantage of the approaches developed in this paper over the current forecasting techniques used by NGN is that, by treating forecasting as a Bayesian statistical problem, we obtain a representation of the uncertainty in the forecasts. This will potentially allow NGN to build their attitude towards risk formally into their decision maing around gas demand.

7 SUMMARY

We have used time series approaches to model the short term hourly and long term daily demand for natural gas in the Northern and North Eastern LDZs. In both cases models were built up using assumptions of normality on the log-demand, seasonal effects and covariates including a weather variable such as temperature. The models capture the main behaviour in the demand for gas well and can be used to improve the operations of NGN.

The short term forecasting model for offtakes represents an opportunity for NGN to move gas around the network more efficiently, reducing the need to redirect gas on an hourly basis to cover unexpected demand levels in specific parts of the network. This would release personnel who are currently needed to do this manually to undertake other work, improving organisational efficiency.

The models could be improved. An extension of the short term forecasting model would be to incorporate explicit temporal dependence in the demand forecasts. That is, when gas demand at a particular offtake has been higher than expected in the previous hour, this would indicate that gas demand at that offtake is likely to be higher than expected in the next hour. This could be built in to the model, for example, using a simple random walk or an AR(1) process as with temperature.

REFERENCES

Congdon, P. (2006). *Bayesian statistical modelling*. Wiley.

National Grid (2012, February). Gas demand forecasting methodology. Technical report.

Rabiner, L. & B. Juang (1986). An introduction to hidden markov models. *IEEE Trans. Acoust. Speech Signal Process. 3*, 4–16.

Shumway, R. & D. Stoffer (2011). *Time Series Analysis and its Applications with R Examples*. New York: Springer.